Exhibit A

nature

INTERNATIONAL WEEKLY JOURNAL OF SCIENCE

Vol 308 No 5957 22-28 March 1984 £1.80 $4.50

FOSSILIZED OCEANIC SULPHIDE DEPOSITS

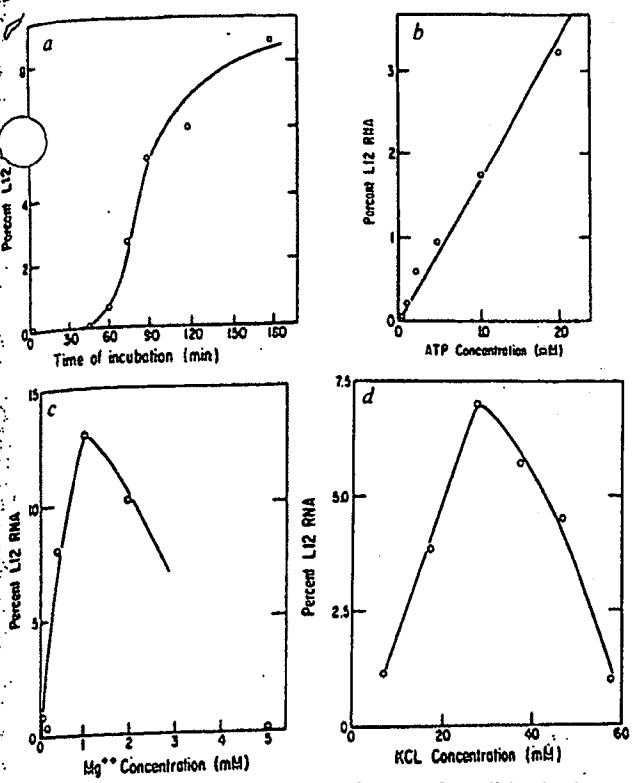

Fig. 2 Time course and cofactor requirements for splicing *in vitro* of an exogenous RNA precursor. *a*, Purified precursor RNA was incubated for various times in the whole cell extract in splicing conditions. The RNA products of the reaction were then analysed by hybridization to M13 pJAW 43 DNA followed by digestion with pancreatic RNase, as described in Fig. 1. The fraction of spliced RNA was determined by densitometric tracing of the autoradiogram. The ATP dependence, *b*, the Mg^{2+} dependence, *c*, and the KCl dependence, *d*, were determined in analogous experiments. RNA precursor synthesized as described in Fig. 1 was purified by Sephadex G-50-150 chromatography to remove the nucleoside triphosphates. The splicing reactions were performed as described in Fig. 1, except for the following modifications. For the experiment shown in *a*, each reaction contained a concentration of Mg^{2+} equal to the ATP concentration plus 0.5 mM. For the experiment of *c*, ATP was present at a concentration of 1 mM. For the experiment of *d*, ATP and Mg^{2+} were present at a concentration of 1 mM and 1.5 mM, respectively.

of incubation. This level of spliced product remained stable for an additional 12-h incubation. The unusually long lag cannot reflect a slow ATP-independent step involving assembly of complexes which then rapidly utilize ATP to cleave and ligate RNA, because preincubation of labelled RNA and extract in the absence of ATP does not shorten the lag period between the addition of ATP and the appearance of spliced product. This suggests that the rate-limiting step, perhaps formation of a complex with substrate RNA, is ATP dependent.

Cofactor requirements for the splicing of mRNA precursors in an extract of HeLa cells are not similar to those of rRNA splicing which does not require ATP[3,4] but are similar to those for the splicing of tRNA precursors; both reactions require ATP and Mg^{2+} (refs 8, 9). The two reactions differ in their Mg^{2+} dependence, however; 10 mM Mg^{2+} is optimal for tRNA splicing[2] while mRNA precursor splicing is inhibited at concentrations of Mg^{2+} above 5 mM. In addition, whereas other nucleoside triphosphates such as UTP, CTP and GTP are able to substitute for the ATP requirement in the tRNA splicing reaction[2], these nucleoside triphosphates are not able to substitute for the ATP requirement in the mRNA splicing reaction (Fig. 1c). These differences may reflect the use of different enzymatic activities by the two types of RNA precursors.

The splicing of tRNA precursors occurs by two distinct and separable steps: endonucleolytic cleavage followed by ligation of the exons[8,9]. Whereas the only cofactor required for the cleavage reaction is Mg^{2+}, both Mg^{2+} and ATP are necessary for the ligation reaction. It is unlikely that a similar division of the mRNA splicing reaction is possible. At present, we have not found that a significant fraction of the exogenously added mRNA precursor is cleaved when incubated in a reaction containing Mg^{2+} but no ATP (S.F.H. and P.J.G., unpublished observations). Also, preincubation of RNA and extract in the absence of ATP does not result in the formation of complexes that quickly process RNA upon addition of the cofactor.

We have previously shown that addition of antisera that react with the U1 small nuclear ribonucleoprotein particles (U1 RNP) inhibited splicing in the coupled transcription/splicing system[10]. Similar experiments using exogenous substrate RNA also show inhibition of splicing by addition of the same anti-U1 RNP sera. The pronounced lag before appearance of spliced product might reflect the slow assembly of an RNP–substrate complex in the reaction mix. Formation of an RNA/RNA duplex by sequences at the 5′ end of U1 RNA and the 5′ splice site has been suggested as an important step in recognition of the mRNA precursor[11,12]. However, there is no inherent reason why this type of recognition should be slow given that the U1 RNA concentration in the reaction is 5 μg ml^{-1}.

P.J.G. acknowledges support from the Helen Hay Whitney Foundation and R.A.P. from the Myron A. Bantrell Charitable Trust. This work was supported by NSF grant PCM-7823230 (currently PCM-8200309), NIH grant P01-CA26717 to P.A.S. and partially from Center for Cancer Biology at MIT (Core) grant NIH-P01-CA14051.

Received 21 October 1983; accepted 6 January 1984.

1. Peebles, C. L., Gegenheimer, P. & Abelson, J. *Cell* 32, 525–536 (1983).
2. Greer, C. L., Peebles, C. L., Gegenheimer, P. & Abelson, J. *Cell* 32, 537–546 (1983).
3. Cech, T. R., Zaug, A. J. & Grabowski, P. J. *Cell* 27, 487–496 (1981).
4. Kruger, K. *et al. Cell* 31, 147–157 (1982).
5. Padgett, R. A., Hardy, S. F. & Sharp, P. A. *Proc. natn. Acad. Sci. U.S.A.* 80, 5230–5234 (1983).
6. Fire, A., Baker, C. C., Manley, J. L., Ziff, E. B. & Sharp, P. A. *J. Virol.* 40, 703 (1981).
7. Manley, J. L., Fire, A., Samuels, M. & Sharp, P. A. *Meth. Enzym.* 101, 568–582 (1983).
8. Peebles, C. L., Ogden, R. C., Knapp, G. & Abelson, J. *Cell* 18, 27–35 (1979).
9. Knapp, G., Ogden, R. C., Peebles, C. L. & Abelson, J. *Cell* 18, 37–45 (1979).
10. Padgett, R. A., Mount, S. M., Steitz, J. A. & Sharp, P. A. *Cell* 35, 101–107 (1983).
11. Lerner, M. R., Boyle, J. A., Mount, S. M., Wolin, S. L. & Steitz, J. A. *Nature* 283, 220–224 (1980).
12. Rogers, J. & Wall, R. *Proc. natn. Acad. Sci. U.S.A.* 77, 1877–1879 (1980).
13. Zain, S. *et al. Cell* 16, 851–861 (1979).
14. Manley, J. L., Fire, A., Cano, A., Sharp, P. A. & Gefter, M. L. *Proc. natn. Acad. Sci. U.S.A.* 77, 3855–3859 (1980).

Similarity of the nucleotide sequences of rat α-lactalbumin and chicken lysozyme genes

Pradman K. Qasba & Surinder K. Safaya

Laboratory of Pathophysiology, National Cancer Institute, Bethesda, Maryland 20205, USA

α-Lactalbumin (α-LA) is a milk protein that interacts with the enzyme galactosyltransferase, modifying its substrate specificity in a way which promotes the transfer of galactose to glucose, resulting in a β−1→4 glycosidic linkage and the synthesis of lactose[1,2]. Lysozyme, an enzyme which catalyses the hydrolysis of a β−1→4 glycosidic linkage in polysaccharides, has been shown to be structurally related to α-LA and it has been proposed that they have arisen from a common ancestral gene[3]. To compare their evolutionary relationships, we report here the complete nucleotide sequence of the rat α-LA gene, including its 5′-flanking sequences, and compare its gene structure with the chicken egg-white lysozyme gene[4]. Both genes contain three introns at similar positions. The first three exons of the two genes have similar nucleotide sequences. The fourth exon of α-LA, which partly codes for the C-terminal residues of the protein, essential for its interaction with galactosyltransferase[5,6], is markedly different from the corresponding exon of the lysozyme gene and is preceded by two (TG)$_n$ repeats.

```
     Eco RI
1    GAATTCCCAT TCCGTGTCCC AAACCTTCCC AGTCCTAGGT AGGATATGAG AGATCCGCAG AGGCAACACA TTTGCTGAAT TTCTGAGCCC GAATCTCTTC
101  CCATTCAGGA GTTTCTGTTC ATCCACCTTG GGTTTCCACA TCCTCCGCAC GAGTTTCAGT CACACTGTGC TATACACTGG TTACCCTGAG CAGAGCGGGC
                      Hha I
201  GGGGTTAAAC AGAGAAGAGC GCCTTTGTCT CTGAAGACTT ACAGATAATA GCCAGTGGGA AGACTTCCAG TTTTTAAGTC TTAGGTGGCT TGGTTAGGAT
          Ava I
301  GGAGACCTCG GGAGAAGTAA AAATCAGGCT TATGAAATGA GCATAGGTAG GTAGGAAGGC AAAGATGACC AAGACCACGA ATGTAGCGTT CCATGACAGC
401  AAATTACAGT AATCACTTGG TACTAGTACC CAGCCTTGAC TTGCTTAGAG GAGATGACTG TAACAGTTTA CCCACGACCC CTTTCCTGTC TGTGGCCAAC
501  CTTGGGGCTA TCAAGAACTA CTTATCCAAA AACAATGTTA CTATGAAATT GCCTTGCATT GACACAATGT CTTAATAACA CCAAGAAGCA TCTCTGGCAG
                                                                 Hind III
601  TGGCGACTAG GGGCAGGGAG CTTGGCAATT TCTTATAGGA CAGCTAGATA TTTTACACAC AAAGCTTTAC CATTATCCTA TCTTCCTCCA TAGGAAGTCT
                                                                                          Fnu 4H1
701  TGACATCCAA GCTGTATTTG GCTCTCATGC CAAATGTCAC CGTGTCCCAT CCCTTTCTCC GAGGCCCTCC GCTATGAGCT GCCGTATCAA GGGGCTTCCT
                                            Ava II
801  GAACTTTGTC CCATGGTTAG GTAGGCCAGA CAAACGCCAT TGGTCCTAGG ATATGATAAA TCCCTTTCCA CATTCTGTAT GTGTTTCGGG GTAGGAATGG
                                                                                                    Fnu 4H1
901  AGTTGGCTGT GGTCAACTGT GTTTTCCAAT CACCTCACTT GTGCTGTCCT GAAGCCTCCT TGTTCCTAAG GTTAGGGACA TCCTGCTTCC TAGAAGCAGC
                             Dde I
1001 CCTGAAAGAA ACAAAGGTTT ACCAAAACTG AGCGAGAAGC TGGATCATTG TAGAACTATA TGGGCAGAAA ACAATACTTG GTATTGACTG GATATAGGAG
1101 GTGGGGAGGG TGGCAGGATG GAGGGAAGTT GGCAGGCTCG GCGTTTCTAT CTTGGCAGAA ACTTCTCGTC ATCATCATCT TCCTAGAGAG AAGGCXTGGT
                                         ▽ ▽ Bam HI
1201 GCTAGGGCCA GAGGCCTTCT TCA[TAAATAAAA]GCAGGTGA AGTGAGTGGG ATCCACATTC AAGGTCTGGG AGCAGGCAAA ATG ATG CGT TTT GTT
                                                                                          MET MET Arg Phe Val
                                                                                          -19             -15
                                     Fnu 4H1
1288 CCT CTG TTC CTA GCG TGT ATT TCG CTG CCT GCC TTT CAA GCC ACA GAG TTT ACA AAA TGT GAG GTG TCC CAC GCC
     Pro Leu Phe Leu Ala Cys Ile Ser Leu Pro Ala Phe Gln Ala Thr Glu Phe Thr Lys Cys Glu Val Ser His Ala
                 -10             -6              -1  1                5                       10
1371 ATT GAA GAC ATG GAT GGC TAT CAA GGC ATC AGC TTG CTT GAA T GTAAGTTTAC ACTTACTTCC TTGACTTCCT CTCATTTCCC AAC
     Ile Glu Asp MET Asp Gly Tyr Gln Gly Ile Ser Leu Leu Glu T
                 15                    20                     25                            Bcl I
1467 TTTCCCTCC TCCTCCTCCT CCTCCTCCTC CTCCTCCTCC TCCTCCTCCT CCTCCTCCTC CTCCTCCTCC TCCGTGGCCA TAGCTCTTCT TGTTCTGTCT T
1557 GATCTCTTC ATTCAACAAG TTCTCCGCCC TTCCTGTTCT CCCATTGTTC TGGCCTTTCT CGACTTTCGA CCAAGATGCT TCTGGACTTT ACCACCTTTGG
1657 AGATTGGTT GGGAAGCCTC CCCCCTGCTC CAAATTAGTT TCTCAATTGT GCCTCTCTGA GACATCTCTG ATGAAAGTTC TTCCTTCTGT CTCCCTCAG
     Ava II                                                                             Fnu 4H1
1755 GG ACC TGT GTT TTA TTC CAC ACC AGT GGC TAT GAC TCA CAA GCT ATC GTC AAG AAC AAT GGC AGC ACA GAG TAT
     rp Thr Cys Val Leu Phe His Thr Ser Gly Tyr Asp Ser Gln Ala Ile Val Lys Asn Asn Gly Ser Thr Glu Tyr
                     30                  35                  40      Ava I       45                  50
1829 GGA CTC TTC CAG ATC AGT AAC AGA AAT TGG TGC AAG AGT AGC GAG TTC CCC GAG TCA GAG AAC ATC TGT GAC ATC
     Gly Leu Phe Gln Ile Ser Asn Arg Asn Trp Cys Lys Ser Ser Glu Phe Pro Glu Ser Glu Asn Ile Cys Asp Ile
                     55                  60                  65                  70                  75
1904 TCC TGT GAC A GTGAGTACCC TCTTTCATCA TGTTCTTCTG TATCCCCATA CCTGCCTCTG AGTCTTCCAA AGCAGGAGGA GTGCTGGGTA CTGTGTG
     Ser Cys Asp L
                                                                          Ava I
2001 ACTTGAGATGGC GGTGCTTAGT AAGAAACTGG TCATGATCTC CGCATGACCC AGAGAGTCCC CTGAGGCTCG GGGAAGGGTA GTTCAAGACA CTTGTGAG
2101 TCTTGAAGTCTG ATACGCTGTA TTTTCAGGGT GTGATGAAGC TGTCGATTTC TTGAGAGGTG TCCTAGGGAA GGGGATGAAG TCCCCAGGAA AGGGAAGC
2201 ATTACTAGATTGGCCATTTCAT ATAGAAAGTC TCTAGGCCAA GCCTTCAGCT CAGAACCAGT CCCCCTGACC GTGGGAGTAG TGGAGCCAGC CCGCTGAG
2301 CTGTATTTCTTAGCTGTCTTGT TCCTTCCTAC ACTCCGCTAG AG TTC CTT GAT GAT GAG CTT GCG GAT GAC ATA GTA TG? GCC AAG
                                                ys Phe Leu Asp Asp Glu Leu Ala Asp Asp Ile Val Cys Ala Lys
                                                   80                  85      Hha I       90
2387 AAG ATC GTG GCT ATC AAA GGG ATC GAC TAC TG GTGAGTCATTG CCTCTGCTCT CCTCTGGAGG CGCTAGCACT TCCCTTTCTA GGCCTT
     Lys Ile Val Ala Ile Lys Gly Ile Asp Tyr Tr
         95                  100
2476 TGAC GGCGTCTCCA AAACAAAATC CATCTAGAGT CCGAAAGGTC TATATTTAAA ATGCTCTTCT GTGTCATTCT TGTGTGACAA CTTTCCCGAT ACAGGA
     Eco RI                                                                                        Bgl II
2576 ATTCCCCTGT TGGAAAGCTG ACTTTCAGAA CTCTCTTTCC TGGGCCAGTT CTCAGCCTTG AAGCACTTGG TGTGGTTCCA CATCTTTTCT AGATCTGCAA
2676 CTTT ATTGATTTTA TCATTATTAT TGTTTTGTTT TTTTAAAGAA GATTTTATTA ATTTATTGTA CTATGTGCGT ACACTGTCAC TGTCTTGAGA CACACC
              Bam HI
2776 AG AAAAGGACAT TGGATCCCAT TACAGGTGGT TGTGAGCCAC CATGTGGTTG CTGGGAGTTG AACCCCAGCA CTTAACCACT GAGCCATCTC TCCAGCCC
2876 TTAT TTTGGTTTTC TGAGGAGGTG TGTGTGTGTG TGTGTGTGTG TGTGTGTGTG TGTGTGTGTG TGTGTAGCAC TGGCTATCTT GGATCTAGCT CTGT
2976 AAGACCAAGC TGGCCTCAAA CTCACAGAGT CTGCTTACTT ATGCCTTCAC AGATTAAAGT CATAGCACCA CCACCACTAG CCCTATATTA TTATTATTAT
3076 TATTATTATT ATTATTATTA TTATTATTGT TATTATTATT TTATTAATAG TGAGCTCCAT TCCACTGAGG AGCCCTTGCC ATCGACCTCA GAATGTCTGT ATT
3176 TCTCTACAAT CTTTCAGCTC ACACTGAGGA TGAACCTAAA TGGAGTTTTA TGCTAGACAT TGTTAATCAA CTGCAACATG AGAGAAAGTG TGTGTGTGTG
3276 TGTGTGTGTG TGTGTGTCTG TGTGAGTGTG TAATAACTGA GCAAAGCACA CGGATATACA TGTATGAAAC TGTCATAATG GAACCCGCTA TTTGGCATAG
3376 TCAAGTTAAAA AGAAAATTCC AGCCAGTCCA TTCGTTTTGT CCATTTCTCC GTGATCAG G AAG GCT CAC AAG CCC ATG TGC TCT GAG AAG
                                                                 p Lys Ala His Lys Pro MET Cys Ser Glu Lys
                                                                   105                 110
                 Hha I           Hpa II                                      Eco RI
3466 CTG GAA CAG TGG CGC TGT GAG AAG CCG GGA GCT CCT GCC CTT GTT GTC CCT GCC CTG AAT TCA GAA ACA CCT GTT
     Leu Glu Gln Trp Arg Cys Glu Lys Pro Gly Ala Pro Ala Leu Val Val Pro Ala Leu Asn Ser Glu Thr Pro Val
     115                 120                 125                 130                 135
3541 CCC TAA GCTACCTCAG TTTGGCTCTT TCTATCGTGA AGATGATCTG TCTCTGAGCC TTGTATCCTG TAGTGACACC ATCGGACTCT AGAGGACTTT TGT
     Pro End
     140
3640 TCAACAGGAG TGCGACTGGC ACACTGGACT GCAAACCCTT GCTTAGTGAC GGCGAGGGTC TTGATGGTGG TTTTACTAAT GCAAGAGAAC CCTCTCCTGT
                                                                                  Fnu 4H1   Ava I
3740 CCCA[AATAAA]GGGCCCAATTTGATGTTCTCTGTTGGTG TTTTCTTTCT ATAGGGCAAA GGGTGCGAAT ACGGGCGGCA GTCCTCGAGA CT 3829
                             ↑
```

Fig. 1 (*left*) Sequence of the sense strand of the rat α-LA gene region. The coding sequences of the exons have been translated and the amino acid residues numbered below. The start of RNA transcription was identified at position 1,247 (▽) by primer extension[10] and at position 1,245 (▽) by S_1 nuclease mapping[11]. The termination site of the RNA transcript at position 3,762 is shown by †. The consensus sequence, TAAATAAAA, and the polyadenylation site are boxed. The short repeat sequences in the 5'-flanking region and unique sequences of the intervening region are underlined.

Methods: A ^{32}P-labelled rat α-LA cDNA clone pα-LA 18[7] was used as a hybridization probe to screen 5×10^6 plaques from the bacteriophage Charon 4A/rat partial *Eco*RI genomic library[8]. Seven individual hybridizing plaques were isolated, plaque purified, and DNA from each recombinant phage was isolated and mapped with different restriction endonucleases. Rat α-LA cDNA has a single *Eco*RI site[7]. The 3' and 5' end fragments to this site were used as hybridization probes. Comparison of the hybridization patterns of the phage DNAs and the restriction enzyme maps showed that there were two classes of recombinant phages. One class contains 6 and 2.7 kb *Eco*RI liver DNA fragments. Only the 6 kb fragment hybridized to the 3' end probe of pα-LA cDNA. The other class contains *Eco*RI fragments of approximately 0.9, 2.5, 1.0, 5.5, 1.1 and 0.5 kb. Only 0.9 and 2.5 kb fragments hybridized to the 5'-end probe of the essentially full-length α-LA cDNA clone, pα-LA 35 (short of 4 nucleotides on the 5' end) or to α-LA mRNA labelled at the 5' end. A recombinant phage carrying overlapping α-LA genomic DNA fragment of the two phage classes could not be identified from this partial *Eco*RI genomic library even after additional screening of 2×10^6 phages. The restriction DNA fragment of the α-LA gene region from rat mammary gland or liver DNA could be arranged into a genomic map indistinguishable from the restriction enzyme map derived from the two recombinant phages. We have recently isolated a recombinant phage carrying an overlapping α-LA rat genomic DNA from a bacteriophage Charon 4A/rat partial *Hae*III genomic library. Rat *Eco*RI DNA fragments from each phage DNA were isolated on 1% low melting agarose gel and subcloned into the *Eco*RI site of pBR322. Plasmid DNA and their *Eco*RI inserts were isolated and purified from such subclones. The linkage of the *Eco*RI fragments of the recombinant phage DNAs was determined by partial digestion of the DNA with *Eco*RI or extensive digestion with a combination of restriction enzymes, followed by Southern blotting[29] and hybridization with the ^{32}P-labelled *Eco*RI inserts of the subclones. The inserts of some of the subclones, pBα-LA 2.5 and pBα-LA 0.9, containing 2.5 and 0.9 kb *Eco*RI liver DNA fragments, respectively, were extensively mapped with restriction endonucleases and sequenced by the Maxam and Gilbert method[30]. The nucleotide sequence from positions 1–2,574 and 2,575–3,522 was determined from pBα-LA 2.5 and pBα-LA 0.9, respectively. The remaining sequence from position 3,523–3,829 was determined from the *Eco*RI/*Bam*HI fragment isolated from the 6 kb genomic rat DNA fragment.

The nucleotide sequence of the entire rat α-LA gene and its 5'-flanking region is shown in Fig. 1. The α-LA gene extends over 2.5 (kb kilobases) of genomic DNA and is composed of four exons and three introns. The nucleotide sequence of the coding region of the genomic DNA is consistent with the previous reported cDNA sequences of pα-LA 18 and pα-LA 17[7] and with the essentially full-length α-LA cDNA clone pα-LA 35 (Fig. 1 legend). The amino acid sequence of the entire rat α-LA protein has been deduced from the DNA sequence of the gene. This has confirmed our previously published partial sequence of rat α-LA[7] which showed that rat α-LA has a 17 residue extension at the carboxy terminus which is proline rich and hydrophobic. The nucleotide sequence of the gene shows that this extension results from a T to G base change in the termination codon. Comparison of the nucleotide deduced amino acid sequence with the amino acid sequence of rat α-LA reported by Prasad *et al.*[9] shows certain differences. We believe that some of these discrepancies are probably a result of an incorrect amino acid assignment made because of deamination of some amino acids during the protein sequence analysis. We cannot explain the remaining discrepancies.

The 5' end of the rat α-LA mRNA was located at G, position 1,247, by the primer extension method[10]. However, it was identified as T at position 1,245 (Fig. 1) by S_1 nuclease mapping[11]. Preceding this initiation site is an A-rich sequence, 5'-TAAATAAA-3', at position 1,224–1,231. This is similar but not identical to the consensus sequence TATAAA[12] found at the initiation site of many other genes. There are also a number of almost identical short repeat sequences preceding this A-rich sequence, shown underlined in Fig. 1. These appear only in the 5' flanking region of this gene. It remains to be established whether these short repeat sequences are in any way related to the induction or inhibition of the α-LA gene expression by multiple hormonal combinations[13]. It is worth noting that a nonanucleotide sequence ATCCCTTTC, which appears at positions 748–756 and 860–868, resembles the first 9 nucleotides of the 19-nucleotide concensus sequence $ATC\frac{CC}{TT}ATT\frac{A}{T}TCTG\frac{G}{T}TTGTA$ thought to be involved in the progesterone receptor recognition site in the ovalbumin gene[14]. The three intervening sequences of the α-LA gene have some unusual features (Fig. 1). In the first intervening sequence a trinucleotide, TCC, is repeated 23 times between positions 1,459 and 1,528. In the third intervening sequence TG dinucleotide is repeated 25 times from position 2,897 to 2,946 and again 22 times from position 3,263 to 3,306. Between these two stretches of TG repeat sequences a trinucleotide, TAT, is repeated from position 2,964 to 3,114. The TG repeats have been shown to occur in other eukaryotic genomes[15,16].

Comparison of rat α-LA and chicken egg white lysozyme gene organization[4] reveals that the three introns are located at similar positions within the coding regions of the two genes (Fig. 2). Rat α-LA exons are 165 (167), 159, 76 and 328 base pairs (bp) long, compared with the corresponding lysozyme exons which are 165, 162, 79 and 180 bp long. The three introns are located on codons 26, 79 and 104 within the coding regions of rat α-LA, compared with codons 28, 82 and 108 of chicken egg white lysozyme. The first and third introns in both the genes interrupt the Trp codon. However the corresponding intervening sequences in the two genes differ in length: in the rat α-LA gene the intervening sequences are 341, 429 and 1,016 bp long and in the chicken egg white lysozyme gene 1,270, 1,810 and 79 bp long. All three introns in both the genes not only begin with GT and end with an AG dinucleotide, sequences thought to be necessary for correct RNA splicing of various other eukaryotic genes[17], but also have additional common sequences at the exon–intron junctions (Fig. 2).

Comparison of the nucleotide sequences of the coding regions of the first three exons of rat α-LA with chicken egg white lysozyme show sequence similarities of 43%, 56% and 46% in exons 1, 2 and 3, respectively. For this comparison, the amino acid sequences of the two proteins in these regions were first aligned and subsequently their nucleotide sequences. This required one additional residue (in the signal peptide) and two

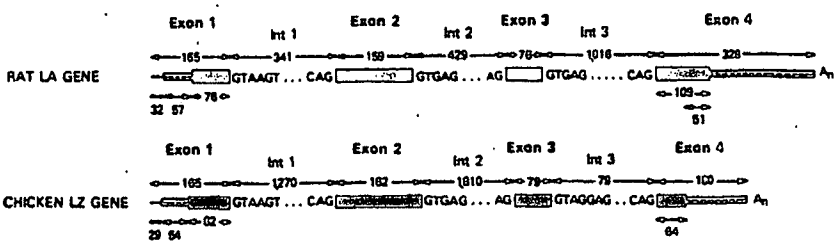

Fig. 2 Comparisons of the rat α-LA and the chicken egg white lysozyme gene regions. Solid boxes represent the regions coding for the structural sequences of the protein; open boxes with a dashed line at the 5'- and at 3'-ends represent, respectively, the region coding the signal peptide of the protein and the 3' noncoding region. In exon 1 the dashed line represents the 5' noncoding region. Numbers represent the bp and Int the intervening sequences.

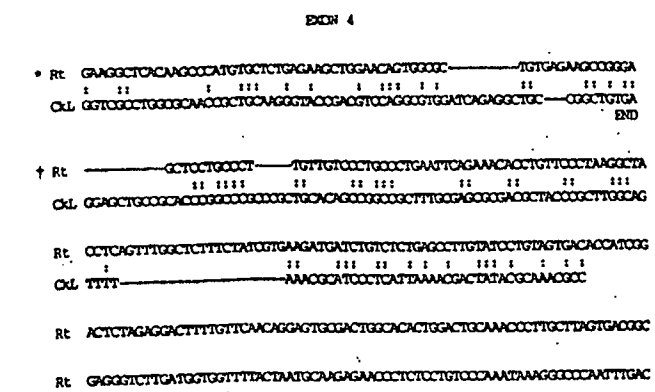

Fig. 3 The nucleotide sequence alignment of the fourth exon of rat α-LA and chicken egg white lysozyme. * Indicates that the alignment of this coding region was based on the amino acid alignment. † Indicates that from this position onwards the sequences were aligned using the NUCALN computer program[23] with the following parameters: k-tuple size, 2; window size, 20; gap penalty, 5.

deletions in the amino acid sequence coded by the first exon, and one deletion each in the second and third exons of α-LA. Using the SEQDP computer program[18] these regions showed statistically significant similarities at the protein level as previously observed with other α-LAs[1]. Exon 2 of α-LA (which has the highest nucleotide and amino acid similarities with exon 2 of lysozyme– codes for amino acid residues considered to be involved in the interaction with galactosyltransferase and partial binding of monosaccharides in the lactose synthase complex[5,6]. Exon 2 of lysozyme codes for residues which carry the catalytic centre and contribute mainly to polysaccharide binding[19,20].

The similarities both at the protein and nucleic acid level, together with the functional similarities of the exons, strengthen the argument that α-LA and lysozyme have evolved from the same primordial gene. However, based on model building, the conformation of the main chain of α-LA, though for the most part similar to lysozyme, differs from it from residue 108 onwards[21]. These residues are encoded by the fourth exon of the two genes which show the least similarity and also differ in length; 328 bp for α-LA and 180 bp for chicken lysozyme (Figs 2, 3). The first 64 nucleotides of the lysozyme exon and 58 nucleotides of α-LA exon code for residues 109–129 and 105–123, respectively, in the two proteins. The 17 residue extension at the carboxy terminus which is unique to the rat α-LA is a result of a single point mutation in the termination codon and has been excluded from this discussion. In this region 5 out of 19 residues of α-LA align with lysozyme only when three deletions and one additional residue are introduced[21] in α-LA. Using the SEQDP[18] and ALIGN[22] programs, the amino acid similarity in this region was not found to be statistically significant. Based on the amino acid alignment we compared the nucleic acid sequence of this region of α-LA and lysozyme (Fig. 3). Using the ALIGN program[22], the two sequences do not show any statistically significant similarities. When the 3′ non-coding regions were aligned using the NUCALN computer program[23], again the nucleotide similarities were not significant (Fig. 3). This, although there is a high degree of similarity implying homology between the other exons, the two genes have diverged in the region of the fourth exon, since this region of the α-LA gene shows no statistically significant similarity with the corresponding lysozyme exon. On the other hand, there is extensive nucleotide similarity between the fourth exon of the rat α-LA gene and the comparable region of human and guinea pig α-LA mRNA.[24] The region coding for residues 105 onwards, which contains amino acids essential for its interaction with galactosyltransferase[5,6], is well conserved among species variants of α-LA. This functionally important region of α-LA is conformationally different from the corresponding region of lysozyme[21], and is coded by a separate nucleotide region preceded by two $(TG)_n$ repeats.

The observed differences in the fourth exon of the two genes suggest that this exon of the primordial gene diverged more rapidly than the other exon through accumulation of point mutations, insertions and deletions, generating a DNA sequence which partly coded for a new functional unit of a protein. Alternatively, according to the split gene hypothesis[25,26], the entire fourth exon might have been replaced by a new region of DNA. TG dinucleotide repeats, present in the intervening sequence preceding the fourth exon of the α-LA gene, have the potential of forming a Z-DNA structure[27]. These sequences have been implicated in genetic recombination, rearrangement and regulation of gene expression[15,28] and may have had some role in generating the observed differences in the fourth exons of the two genes. Thus, α-LA and lysozyme are two related proteins with different functions whose critical divergence appears to be concentrated in one exon and may therefore serve as a useful model for the study of the evolution of new proteins.

We thank Drs David Lipman and Jacob Maizel for carrying out the computer analyses and for their valuable suggestions on the manuscript.

Received 14 September 1983; accepted 5 January 1984.

1. Brew, K. & Hill, R. L. *Rev. Physiol. biochem. Pharmac.* **72**, 105–157 (1975).
2. Hill, R. L. & Brew, K. *Adv. Enzym. Related Areas molec. Biol.* **43**, 411–490 (1975).
3. Brew, K., Vanaman, T. C. & Hill, R. L. *J. biol. Chem.* **242**, 3747–3749 (1967).
4. Jung, A., Sippel, A. E., Grez, M. & Shutz, G. *Proc. natn. Acad. Sci. U.S.A.* **77**, 5759–5763 (1980).
5. Sinha, S. K. & Brew, K. *J. biol. Chem.* **256**, 4193–4204 (1981).
6. Sinha, S. K. & Brew, K. in *Proteins in Biology & Medicine* (eds Bradshaw, R. A. *et al.*) 379–393 (Academic, New York, 1982).
7. Dandekar, A. M. & Qasba, P. K. *Proc. natn. Acad. Sci. U.S.A.* **78**, 4853–4857 (1981).
8. Sargent, T. D. *et al. Proc. natn. Acad. Sci. U.S.A.* **76**, 3256–3260 (1979).
9. Prasad, R. V., Butkowski, R. J., Hamilton, J. W. & Ebener, K. E. *Biochemistry* **21**, 1479–1482 (1982).
10. Ghosh, P. K. *et al. J. molec. Biol.* **126**, 813–846 (1978).
11. Weaver, R. F. & Weissmann, C. *Nucleic Acids Res.* **6**, 1175–1193 (1979).
12. Goldberg, M. thesis, Stanford Univ. (1979).
13. Topper, Y. J. & Freeman, C. S. *Physiol. Rev.* **60**, 1049–1106 (1980).
14. Mulvihill, R. E., LePennec, J-P. & Chambon, P. *Cell* **24**, 621–632 (1982).
15. Nordheim, A. *et al. Nature* **294**, 417–422 (1981).
16. Hamada, H. & Petrino, M. G. & Kakunaga, T. *Proc. natn. Acad. Sci. U.S.A.* **79**, 6465–6469 (1982).
17. Breathnach, R. *et al. Proc. natn. Acad. Sci. U.S.A.* **75**, 4853–4857 (1978).
18. Goad, W. B. & Kanehisa, M. I. *Nucleic Acids Res.* **10**, 247–263 (1982).
19. Kelly, J. A., Sielecki, A. R., Sykes, B. B., James, M. N. & Phillips, D. C. *Nature* **282**, 875–878 (1979).
20. Imoto, T., Johnson, L. N., North, A. C. T., Phillips, D. C. & Rupley, J. A. in *The Enzymes* Vol. 7 (ed. Boyer, P. D.) 665–868 (Academic, New York, 1972).
21. Browne, W. J. *et al. J. molec. Biol.* **42**, 65–86 (1969).
22. Dayhoff, M. O. in *Atlas of Protein Sequence and Structure* Vol. 5, Suppl. 3 (ed. Dayhoff, M. O.) 1–8 (National Biomedical Research Foundation, Washington, DC, 1978).
23. Wilber, W. G. & Lipman, D. J. *Proc. natn. Acad. Sci. U.S.A.* **80**, 726–731 (1983).
24. Hall, L., Craig, R. K., Edbrooke, M. R. & Campbell, P. N. *Nucleic Acids Res.* **10**, 3503–3515 (1982).
25. Gilbert, W. *Nature* **271**, 501 (1978).
26. Tonegawa, S., Maxam, A. M., Tizard, R., Bernard, O. & Gilbert, W. *Proc. natn. Acad. Sci. U.S.A.* **75**, 1485–1489 (1978).
27. Arnott, S., Chandrasekaran, R., Birdsall, D. L., Leslie, A. G. W. & Ratcliff, R. L. *Nature* **293**, 743–745 (1980).
28. Nordheim, A. & Rich, A. *Nature* **303**, 674–679 (1983).
29. Southern, E. M. *J. molec. Biol.* **98**, 503–517 (1975).
30. Maxam, A. M. & Gilbert, W. *Proc. natn. Acad. Sci. U.S.A.* **74**, 560–564 (1971).

Fibronectin receptors on *Trypanosoma cruzi* trypomastigotes and their biological function

M. A. Ouaissi, D. Afchain, A. Capron & J. A. Grimaud*

Centre d'Immunologie et de Biologie Parasitaire, INSERM U 167, CNRS, ERA 422, Institut Pasteur, 15 rue Camille Guérin, 59019 Lille Cedex, France

* Laboratoire de Pathologie Cellulaire du Foie, CNRS, ERA 819, Institut Pasteur, 77 rue Pasteur, 69365 Lyon, France

Successful invasion of mammalian cells by pathogenic parasites is generally considered, from circumstantial evidence, to be a consequence of specific mechanisms of recognition of cell surface components[1–3]—this has stimulated investigations of the biochemical characterization of such molecules[4–6]. Several studies of trypanosomiasis have examined the ability of parasites to interact with mammalian cells[7–9]. However, knowledge of